

PROBABILISTIC ROBUSTNESS ANALYSIS — RISKS, COMPLEXITY AND ALGORITHMS

XINJIA CHEN, KEMIN ZHOU AND JORGE L. ARAVENA

ABSTRACT. It is becoming increasingly apparent that probabilistic approaches can overcome conservatism and computational complexity of the classical worst-case deterministic framework and may lead to designs that are actually safer. In this paper we argue that a comprehensive probabilistic robustness analysis requires a detailed evaluation of the robustness function and we show that such evaluation can be performed with essentially any desired accuracy and confidence using algorithms with complexity linear in the dimension of the uncertainty space. Moreover, we show that the average memory requirements of such algorithms are absolutely bounded and well within the capabilities of today's computers.

In addition to efficiency, our approach permits control over statistical sampling error and the error due to discretization of the uncertainty radius. For a specific level of tolerance of the discretization error, our techniques provide an efficiency improvement upon conventional methods which is inversely proportional to the accuracy level; i.e., our algorithms get better as the demands for accuracy increase.

1. INTRODUCTION

In recent years, a number of researchers have proposed probabilistic control methods for overcoming the computational complexity and conservatism of the deterministic worst-case robust control framework (e.g., [1]–[7], [11]–[20] and the references therein).

The philosophy of probabilistic control theory is to sacrifice cases of extreme uncertainty. Such paradigm has lead to the concept of *confidence degradation function* (originated by Barmish, Lagoa and Tempo [2]), which has demonstrated to be extremely powerful for the robustness analysis of uncertain systems. Such function, $\mathcal{P}(\cdot)$, is defined as $\mathcal{P}(r) = \inf_{0 < \rho \leq r} \mathbb{P}(\rho)$ with

$$\mathbb{P}(\rho) = \text{vol}\{X \in \mathcal{B}_\rho \mid \text{The robustness requirement is guaranteed for } X\} / \text{vol}\{\mathcal{B}_\rho\}$$

where the volume function $\text{vol}\{\cdot\}$ is the Lebesgue measure, and \mathcal{B}_ρ denotes the uncertainty bounding set with radius ρ . Interestingly, it was discovered in [2] that such function is not necessarily monotone decreasing in the uncertainty radius. In view of this fact and for the purpose of avoiding the confusion with the concept of *confidence band*, used in the evaluation of the accuracy of the estimate of $\mathbb{P}(r)$, the confidence degradation function is referred to as the *robustness function* in this

Date: August 2006.

Key words and phrases. Robustness analysis, risk analysis, randomized algorithms, uncertain system, computational complexity.

This research was supported in part by grants from NASA (NCC5-573), LEQSF (NASA/LEQSF(2001-04)-01), the NNSFC Young Investigator Award for Overseas Collaborative Research (60328304) and a NNSFC grant (10377004).

paper. Accordingly, a graph representation of the robustness function is called the *robustness curve*. It can be seen that the robustness function is a natural extension of the concept of robustness margin. From the robustness curve, one can determine the *probabilistic robustness margin* [2] and estimate the deterministic robustness margin.

In addition to overcoming the NP hard complexity and conservatism of deterministic robustness analysis methods, the robustness function can address very complex problems which are intractable by deterministic worst-case methods. Moreover, the probability that the robustness requirement is guaranteed can be inferred from the robustness function, while the deterministic margin loses the connection with such probability. Based on the assumption that the density function of uncertainty is radially symmetric and non-increasing with respect to the norm of uncertainty, it has been shown in [2] that the probability that the robustness requirement is guaranteed is no less than $\mathcal{P}(r) = \inf_{\rho \in (0, r]} \mathbb{P}(\rho)$ when the uncertainty is included in a bounding set with radius r . The underlying assumption is supported by modeling and manufacturing considerations that the uncertainty is unstructured so that all directions are equally likely and that small perturbations from the nominal model are more likely than large perturbations. Since $\mathbb{P}(\cdot)$ is not monotonically decreasing [2], the lower bound of the probability depends on $\mathbb{P}(\rho)$ for all $\rho \in (0, r]$. It is not clear whether it is feasible to estimate $\mathcal{P}(r)$ since the estimation of $\mathbb{P}(\rho)$ for every ρ relies on intensive Monte Carlo simulation and $\mathbb{P}(\rho)$ needs to be estimated for numerous values of ρ . For such probabilistic method to overcome the NP hard of worst-case methods, it is necessary to show that the complexity for estimating $\mathcal{P}(r)$ for a given r is polynomial in terms of computer running *time* and memory *space*. In this paper, we demonstrate that the complexity in terms of space and time is surprisingly low and is *linear in the uncertainty dimension and the logarithm of the relative width of the range of uncertainty radius*.

In the next section we argue that both the deterministic robustness margin and its risk-adjusted version – the probabilistic robustness margin have inherent limitations. We address those limitations through the use of the robustness function that can describe the performance of a system over a wide range of uncertainties. In order to construct the robustness function for wide range of uncertainty radii, the conventional method independently estimate $\mathbb{P}(r_i)$ for each grid points of uncertainty. If there are m grid points and N is the sample size for each radius, then the total number of simulations is Nm . In Section 3, we use the sample reuse principle and demonstrate that the robustness curve for arbitrarily wide range of uncertainty radii can be accurately constructed with surprisingly low complexity. Clearly, the number of grid points, m , must tend to infinity as the tolerance tends to zero. However, we show that with our algorithms, the *equivalent number of grid points* (ENGP), m_{eq} , is strictly bounded from above in the sense that in order to guarantee the same level of accuracy for the estimation of the robustness function, the required average computational effort is the same as that of a conventional grid with m_{eq} points. Moreover, we show that the average memory requirement is also absolutely bounded and is well within the reach of modern computers.

The remainder of the paper is organized as follows. Section 2 provides an example illustrating the pitfalls of deterministic robustness margin and the probabilistic robustness margin. Section 4 discusses the control of estimation error of the robustness function and the required complexity. Section 5 investigates the difficulties of

the conventional data structure. Section 6 describes our new algorithms, analyzes the complexity of data processing and memory space, and introduces the concept of confidence band. The proofs of all the theorems are included in the Appendices.

2. THE RISK OF ROBUSTNESS MARGINS

In this section we make the case for the need to have a robustness function in order to properly estimate how well a control system tolerates uncertainties. Conventional robust control approaches the issue with a “worst case” philosophy. In this regard, it has been demonstrated (Chen, Aravena and Zhou, [5]) that it is not uncommon for a probabilistic controller to be significantly less risky than a deterministic worst-case control. The reasons are the “uncertainty in modeling uncertainties” and the fact that the worst-case design cannot, in some instances, be “all encompassing.” Therefore, the worst-case approach has an associated risk that usually is overlooked, while the probabilistic approach acknowledges the risk and manages it.

From manufacturing and modeling considerations, it is sensible to assume that the density of the distribution of uncertainty decreases with increasing uncertainty norm. Such assumption leads to the worst-case property of uniform distribution in robustness analysis [2]. However, the decay rate of density is generally unknown to the designer. Therefore, for a given uncertainty radius r , one does not have good knowledge about the coverage probability of the uncertainty set \mathcal{B}_r . It is important to note that the system robustness depends critically on the distribution of uncertainty norm.

Attempts to improve the analysis have led to the definitions of a *deterministic robustness margin* and a *probabilistic robustness margin*. Both are numbers that purportedly allow the user to estimate the tolerance to uncertainties. We contend that both can be misleading, and for essentially the same reason. To demonstrate this view point, we consider a feedback system shown in Figure 1.

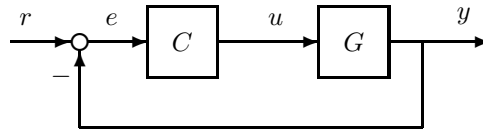


FIGURE 1. Standard Feedback Configuration

The transfer function of the plant is $G(s) = \frac{q}{s-p}$ where p and q are uncertain parameters. The uncertainty bounding set with radius $r > 0$ is

$$\mathcal{B}_r = \{(x, y) : |x - q_0| \leq r, \quad |y - p_0| \leq r\}, \quad p_0 < 0, \quad q_0 > 0.$$

Consider two controllers $C_A = \frac{K_A}{s+\sigma}$, $\sigma > 0$ and $C_B = K_B$ such that

$$1 < K_B < \frac{K_A}{\sigma}, \quad \frac{K_A q_0 - \sigma p_0}{K_A + \sigma} < \sigma - p_0.$$

Suppose that the robustness requirement is stability. It can be shown that the robustness function for controller A is

$$\mathcal{P}^A(r) = \begin{cases} 1 & \text{for } 0 < r < \rho_A; \\ \frac{1}{2} - \frac{K_A \left(r + \frac{\sigma}{K_A} - q_0 \right)^2}{8\sigma r^2} - \frac{p_0 - \beta}{2r} & \text{for } \rho_A \leq r \leq \rho_A^*; \\ \frac{1}{2} - \frac{(r + \beta - p_0) \left(r + \frac{\sigma(\beta + p_0 - r)}{2K_A} - q_0 \right)}{4r^2} - \frac{(p_0 - \beta)}{2r} & \text{for } r > \rho_A^* \end{cases}$$

where

$$\rho_A = \frac{K_A q_0 - \sigma p_0}{K_A + \sigma}$$

is the deterministic robustness margin, $\beta = \min(\sigma, p_0 + r)$, and $\rho_A^* = \frac{K_A q_0 - \sigma p_0}{K_A - \sigma}$. It can be shown that the robustness function for controller B is given by

$$\mathcal{P}^B(r) = \begin{cases} 1 & \text{for } 0 < r < \rho_B; \\ 1 - \frac{K_B \left(r + \frac{p_0 + r}{K_B} - q_0 \right)^2}{8r^2} & \text{for } \rho_B \leq r \leq \rho_B^*; \\ \frac{1}{2} - \frac{\frac{p_0}{K_B} - q_0}{2r} & \text{for } r > \rho_B^* \end{cases}$$

where

$$\rho_B = \frac{K_B q_0 - p_0}{K_B + 1}$$

is the deterministic robustness margin and $\rho_B^* = \frac{K_B q_0 - p_0}{K_B - 1}$.

We consider an example with $p_0 = -10$, $q_0 = 50$, $\sigma = 40$, $K_A = 100\sigma$, $K_B = 10$. The corresponding robustness functions are displayed in Figure 2. We obtained deterministic margins $\rho_A = 49.6040$, $\rho_B = 46.3636$. Since $\rho_A > \rho_B$, a comparison based on the deterministic margin simply suggests that controller A is more robust than controller B . Quite contrary, a judgement based on the robustness curves indicates that controller B may be more robust. The risk of the probabilistic robustness margin can also be illustrated by this example.

Robust analysis should be able to help a designer to reliably determine which controller design is more robust. However, it appears that the concepts of robustness margin fail to meet such fundamental needs of control engineering. On the other hand, the robustness curve serves the purpose of giving the designer complete information on how well a control system tolerates uncertainties.

From the previous discussion, it can be seen that there are two crucial factors to be considered in order to make a reliable judgment about the system robustness:

- (i): How fast the robustness curve rolls off.
- (ii): The dependency of coverage probability of uncertainty bounding set \mathcal{B}_r on the radius r .

The second factor can be a difficulty since a designer generally lacks knowledge of the coverage probability corresponding to a bounding set of fixed radius. To overcome such difficulty, the only choice is to construct the robustness curve for a wide range of uncertainty radius. The construction of the robustness curve may be seen as a computationally challenging task since the probability of guaranteeing robustness requirement needs to be estimated for many values of uncertainty radius. However, as we demonstrate in the next section, using the sample reuse principle one can construct the robustness curve for virtually the entire scope of uncertainty range $(0, \infty)$ with absolutely bounded average computational requirements, regardless of

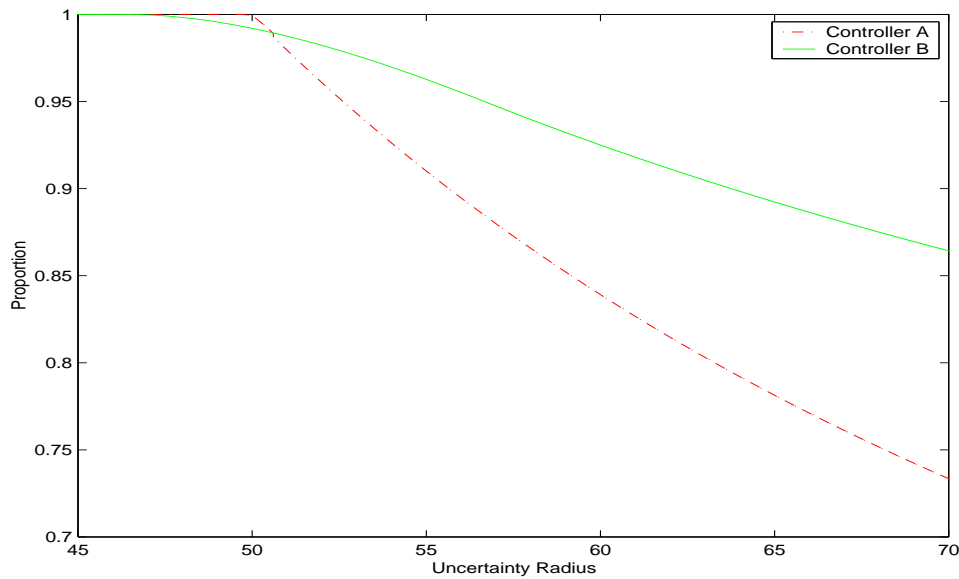


FIGURE 2. Comparison of Controller Alternatives.

the size of the grid. For example, we shall show that for an uncertainty range as large as $(10^{-10}, 10^{10})$, in the average, one needs less than 50 times memory and computational resources than those needed to evaluate the uncertainty range $(1, e)$ with the same resolution.

3. EQUIVALENT NUMBER OF GRID POINTS

Throughout this paper, we assume that the uncertainty sets are homogeneous star-shaped (e.g., [2]). That is, the uncertainty bounding set with radius r is $\mathcal{B}_r = \{rX \mid X \in \mathcal{B}_1\}$ where \mathcal{B}_1 denotes the uncertainty bounding set such that $cX \in \mathcal{B}_1$ for any $X \in \mathcal{B}_1$ and any $c \in [0, 1]$. Clearly, most of the commonly used uncertainty bounding sets such as the l_p balls and spectral norm balls are homogeneous star-shaped.

We shall consider the problem of constructing the robustness curve for arbitrary robustness requirement \mathbf{P} under such assumption of uncertainty sets. Conventionally, the robustness curve for a range of uncertainty radii $[\frac{a}{\lambda}, a]$ with $a > 0$, $\lambda > 1$ is constructed by choosing a set of grid points $\frac{a}{\lambda} = r_1 < r_2 < \dots < r_m = a$ and, for every grid point, performing N i.i.d. Monte Carlo simulations. Hence, the total number of simulations is a deterministic constant mN . To reduce computational complexity, we shall make use of the following intuitive concept:

Let X be an observation of a random variable with uniform distribution over $\mathcal{B}_\rho \supseteq \mathcal{B}_r$ such that $X \in \mathcal{B}_r$. Then X can also be viewed as an observation of a random variable with uniform distribution over \mathcal{B}_r .

In order to apply such concept, it is necessary to perform the simulation in a backward direction so that appropriate evaluations of the robust requirement for larger uncertainty sets can be saved for the use of later simulations on smaller

uncertainty sets [6]. The sample reuse principle allows a single simulation to be used for multiple radii. Thus, the actual total number of simulations is significantly reduced. In order to quantify this reduction we introduce the *equivalent number of grid points* (ENGP), m_{eq} , defined as

$$m_{eq} = \frac{\text{expected total number of simulations}}{N}.$$

In our approach, the number of simulations required at uncertainty radius r_i , denoted by \mathbf{n}_i for $i = 1, \dots, m$, is a random number. The total number of simulations can be represented by the random variable $\mathbf{n} = \sum_{i=1}^m \mathbf{n}_i$. The expected value of the total number of simulations is $\mathbb{E}[\mathbf{n}] = \sum_{i=1}^m \mathbb{E}[\mathbf{n}_i]$ where $\mathbb{E}[X]$ denotes the expectation of random variable X . Hence, we can formally define

$$m_{eq} = \frac{\mathbb{E}[\mathbf{n}]}{N}.$$

Due to sample reuse, we can achieve a substantial reduction of simulations, i.e., $\mathbb{E}[\mathbf{n}] \ll mN$. To quantify the reduction of the computational effort, we have introduced the notion of *sample reuse factor* [6], which is defined as

$$(3.1) \quad \mathcal{F}_{\text{reuse}} \stackrel{\text{def}}{=} \frac{mN}{\mathbb{E}[\mathbf{n}]} = \frac{m}{m_{eq}}.$$

In our approach, N i.i.d simulation results are collected for each grid point. Hence, the accuracy of estimation is the same as that of the conventional method. However, the average number of simulations in our approach is $\mathbb{E}[\mathbf{n}]$, which is equivalent to the complexity of m_{eq} grid points in the conventional scheme. As a direct consequence of Theorem 1 of [6], we have that, for any discretization scheme, m_{eq} is independent of the sample size N . Moreover, we have the following general results.

Theorem 1. *Let d be the dimension of uncertainty parameter space. Then, for arbitrary gridding scheme, the equivalent number of grid points based on the principle of sample reuse is strictly bounded from above by $1 + d \ln \lambda$, i.e.,*

$$m_{eq} < 1 + d \ln \lambda.$$

See Appendix A for a proof. By an “arbitrary” discretization scheme, we mean two things: i) the number of grid points can be arbitrarily large; ii) the grid points can be distributed arbitrarily over the specified range of uncertainty radius.

A fundamental question of robust control is whether randomized algorithms have polynomial complexity. In light of the fact the cost of each simulation depends on problem cases, the computational complexity is usually measured in terms of the number of simulations. This theorem reveals the following important facts:

- (a): The complexity is linear in the dimension of the uncertainty space. Thus our algorithms overcome the curse of dimensionality.
- (b): The complexity depends linearly in the logarithm of the “relative” width, λ , of the interval of uncertainty radii. This proves that our algorithms are capable of estimating the robustness function for a wide range of uncertainty.
- (c): Our algorithms can arbitrarily reduce the grid error, while keeping the complexity strictly below a constant bound.

In order to illustrate these points, Figure 3 displays the variation of m_{eq} for various dimensions of the uncertainty space and for values of λ up to $\lambda = 10^{20}$ corresponding to the uncertainty range $(10^{-10}, 10^{10})$ (which may be deemed a good approximation to $(0, \infty)$). Notice that even for dimensions as high as $d = 1024$ the equivalent number of grid points, m_{eq} , is very reasonable.

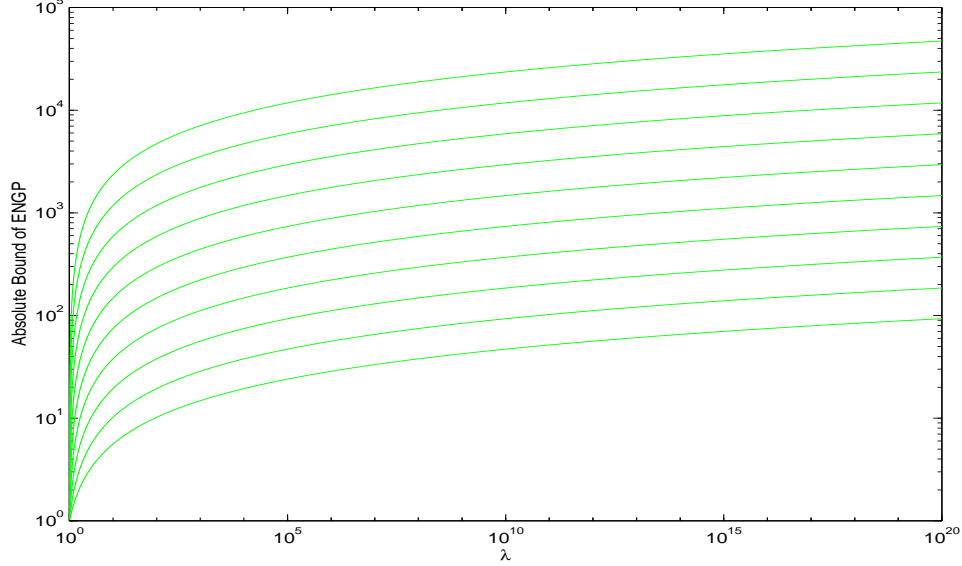


FIGURE 3. Absolute Bounds for m_{eq} (ENGP) ($d = 2^i$, $i = 1, \dots, 10$).

Finally in this section, we consider the case where we need to estimate $\mathbb{P}(r)$ for $r \in [\gamma, U]$ where $\gamma > 0$ is a constant, and U is an estimate of the probabilistic robustness margin calculated by randomized algorithms. Clearly, U is a random variable. If U depends on samples which are independent of the samples generated from the uncertainty set with radius $r \in [\gamma, U]$ we have the following result:

For any gridding scheme,

$$(3.2) \quad m_{eq} < 1 + d \ln \frac{\mathbb{E}[U]}{\gamma}.$$

To prove (3.2), notice that $\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]$ for any random variables X and Y . Hence, by Theorem 1,

$$m_{eq} = \mathbb{E}[\mathbb{E}[m_{eq} | U]] < \mathbb{E} \left[1 + d \ln \frac{U}{\gamma} \right] = 1 + d \mathbb{E} \left[\ln \frac{U}{\gamma} \right] < 1 + d \ln \frac{\mathbb{E}[U]}{\gamma}$$

where the last inequality is obtained from applying the Jensen's inequality to the concave function $\ln(\cdot)$.

4. ERROR CONTROL

In addition to efficiency, another important issue in any numerical approach is error control. This point has been emphasized in many control engineering problems. For instance, when computing the H_∞ norm of a system, a lower bound and an upper bound are obtained and is required that the gap between them be less than a prescribed tolerance. A similar situation arises in the computation of the structured singular value (μ).

For the specific case of the estimation of the robustness function, there are two sources of error: i) the statistical sampling error due to the finiteness of the sample size, N (sample size error); ii) the discretization error due to the finite number of points in any partition. Control of the sample size error has been well studied and emphasized. Existing techniques include the Chernoff bounds [8], binomial confidence interval [7, 9], etc. However, we claim that control of discretization error is not sufficiently emphasized. In fact, one can argue that *controlling the sample size error can be meaningless if the discretization error is not controlled*. This will be the case, for example, for those situations where a risk at the level of a small ε (e.g., $\varepsilon = 0.001$) may be significant or unacceptable. *How can any estimation be useful if the discretization error is not ensured to be less than the tolerance ε ?*

In this section, we first introduce an interpolation result necessary to analyze error control methods. Afterward, we discuss two different schemes which insure a discretization error less than a given $\epsilon \in (0, 1)$. The first is a uniform partition whereby the uncertainty radius interval $[\frac{a}{\lambda}, a]$ is partitioned by m points

$$(4.1) \quad r_i = a - \frac{(m-i)(\lambda-1)}{(m-1)\lambda}a, \quad i = 1, \dots, m.$$

In the second scheme we consider a geometric type partition of the form

$$(4.2) \quad r_i = a \left(\frac{1}{\lambda} \right)^{\frac{m-i}{m-1}}, \quad i = 1, \dots, m.$$

For any partition of the uncertainty radius interval, we have the following linear interpolation results.

Theorem 2. *Given an arbitrary partition of the uncertainty radius interval $[\frac{a}{\lambda}, a]$ with $\frac{a}{\lambda} = r_1 < r_2 < \dots < r_m = a$, define*

$$\mathbb{P}^*(r) = \frac{(r - r_i) \mathbb{P}(r_{i+1}) + (r_{i+1} - r) \mathbb{P}(r_i)}{r_{i+1} - r_i},$$

$$g(r) = (r_{i+1} - r) \left(\frac{r}{r_i} \right)^{-d} + (r - r_i) \left(\frac{r_{i+1}}{r} \right)^{-d}.$$

Then, for all $r \in [r_i, r_{i+1}]$,

$$|\mathbb{P}(r) - \mathbb{P}^*(r)| \leq 1 - \frac{g(r_*)}{r_{i+1} - r_i} \leq \frac{d}{2r_i}(r_{i+1} - r_i)$$

where $r_* \in (r_i, r_{i+1})$ is the unique solution of equation

$$\left(\frac{r_{i+1}}{r} \right)^{-d} \left[1 + \left(1 - \frac{r_i}{r} \right) d \right] - \left(\frac{r}{r_i} \right)^{-d} \left[1 + \left(\frac{r_{i+1}}{r} - 1 \right) d \right] = 0$$

with respect to r , which can be solved by a bisection search.

See Appendix B for a proof. As mentioned before, these interpolation results will be used in the construction of a tight confidence band for the robustness function.

Remark 1. To guarantee a prescribed tolerance $\epsilon \in (0, 1)$, the number of grid points must be larger than a certain number. It has been shown by Barmish, Lagoa and Tempo [2] that if

$$(4.3) \quad m \geq 1 + \frac{2(\lambda - 1)d}{\epsilon}$$

then $|\mathbb{P}(r) - \mathbb{P}(r_i)| < \epsilon \quad \forall r \in [r_i, r_{i+1}]$ for $i = 1, \dots, m-1$. This bound shows that, for fixed error ϵ , the complexity is polynomial. From another perspective, it also shows that the number of grid points and computational complexity tend to infinity as the tolerance tends to zero. For example, the robustness analysis problem for complex uncertainty of size 30×30 over an interval of uncertainty with $\lambda = 10$, requires $m \geq 3,240,000,001$ in order to guarantee $\epsilon \leq 10^{-5}$. The bound, however, does not account for the sample reuse principle. Using our approach the equivalent number of grid points for this case is bounded from above by $1 + 1800 \times \ln(10)$.

The following result is our extension of the result by Barmish et al., cited above, and quantifies the advantage of using linear interpolation.

Theorem 3. *Let*

$$(4.4) \quad m = 2 + \left\lfloor \frac{(\lambda - 1)d}{2\epsilon} \right\rfloor$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Then, for a uniform gridding scheme,

$$|\mathbb{P}(r) - \mathbb{P}^*(r)| < \epsilon \quad \forall r \in [r_i, r_{i+1}]$$

for $i = 1, \dots, m-1$. Moreover, the equivalent number of grid points is

$$m_{\text{eq}}(\epsilon) = m - \sum_{i=1}^{m-1} \left(1 - \frac{1}{\frac{m-1}{\lambda-1} + i} \right)^d.$$

See Appendix C for a proof.

Remark 2. We point out that when using linear interpolation the number of grid points given by (4.4) is approximately $\frac{1}{4}$ of the bound given by (4.3).

We now analyze a discretization scheme whereby the partition of the uncertainty interval under study is defined by a geometric series.

Theorem 4. *For a geometric discretization scheme with*

$$m = 2 + \left\lfloor \frac{\ln \lambda}{\ln \left(1 + \frac{2\epsilon}{d} \right)} \right\rfloor$$

and

$$r_i = a \left(\frac{1}{\lambda} \right)^{\frac{m-i}{m-1}}$$

for $i = 1, \dots, m$, the following statements hold true:

(I):

$$|\mathbb{P}(r) - \mathbb{P}^*(r)| < \epsilon \quad \forall r \in [r_i, r_{i+1}], \quad i = 1, \dots, m-1.$$

(II):

$$m_{\text{eq}}(\epsilon) = 1 + \left(1 + \left\lfloor \frac{\ln \lambda}{\ln \left(1 + \frac{2\epsilon}{d} \right)} \right\rfloor \right) \left[1 - \left(\frac{1}{\lambda} \right)^{1 + \left\lfloor \frac{d}{\ln \left(1 + \frac{2\epsilon}{d} \right)} \right\rfloor} \right].$$

(III):

$$\mathcal{F}_{\text{reuse}} > \frac{1}{2\epsilon} \left(1 - \frac{1}{1 + d \ln \lambda} \right).$$

See Appendix D for a proof.

Remark 3. Since $1 + d \ln \lambda \gg 1$ in many situations, the sample reuse factor for the geometric discretization scheme may be written in a more elegant form. That is,

$$\mathcal{F}_{\text{reuse}} \approx \frac{1}{2\epsilon}$$

which is inversely proportional to the tolerance of the discretization error. For example, to ensure that the discretization error is less than 10^{-4} , which is a rather weak requirement for many applications, our algorithm reduces the computational effort by a factor of 5,000 when compared to a conventional approach.

The two discretization schemes considered here, and others, have bounded complexity, but the distributions of the total number of simulations are different. Hence it is reasonable to ask if there is a “best discretization.” Our results indicate that the geometric scheme is generally more efficient, as shown by the comparison of grid points in Figure 4 and the comparison of ENGP in Figure 5.

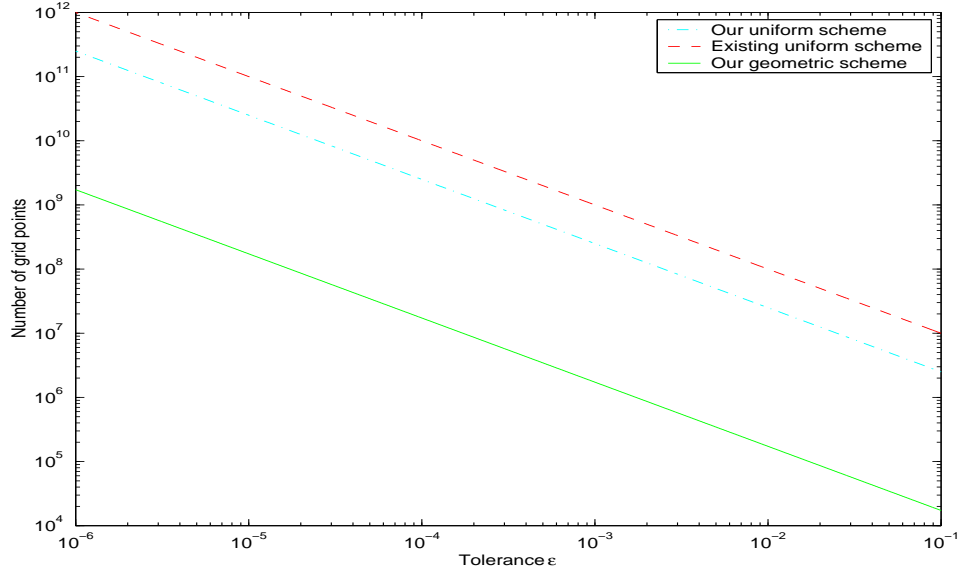


FIGURE 4. Comparison of Number of Grid Points ($\lambda = 10^3$, $d = 500$)

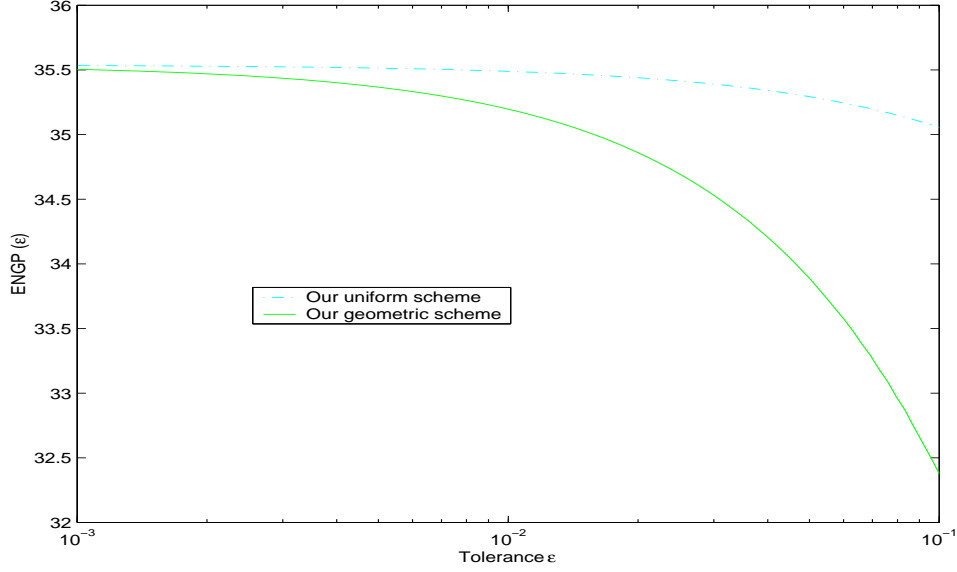


FIGURE 5. Comparison of m_{eq} ($\lambda = 10^3$, $d = 5$, $1 + d \ln \lambda = 35.5388$)

5. THE DIFFICULTIES OF CONVENTIONAL DATA STRUCTURE

Our previous sample reuse algorithm [6] uses the same data structure as that of the conventional algorithm. That is, the data structure for implementing the algorithms is basically a matrix of fixed size. In such data structure, for each grid point r_i , there is a record (k_i, n_i) where k_i represents the number of cases guaranteeing (or violating) the robustness requirement among n_i simulations. In the course of experiment, the number n_i is increment from 0 to sample size N . In the following two subsections, we demonstrate that the conventional data structure is not suitable for controlling the error due to finite gridding.

5.1. The Issue of Data Processing. Clearly, the total number of records is exactly the number of grid points m . For the conventional method, to accomplish N simulations for each grid points, the total number of updating the data record is Nm . As illustrated in Section 4, to control the error due to finite gridding requires an extremely large number, m , of grid points even for moderate requirement of ϵ . Therefore, Nm is usually a *very large number*. It can be shown that *if the sample reuse algorithm employs the same data structure as that of the conventional method, then, for any gridding scheme with m grid points, the total number of times of updating the data record is also Nm* . This is true because, for every time a record (k_i, n_i) is updated, the number n_i can only be increased by 1, and the number n_i must be N when the experiment is completed. To have a feeling that the data processing with the conventional data structure is a severe challenge, one can consider the example discussed in Remark 1 of Section 4. With $m \geq 3,240,000,001$ and normal sample size $10^4 < N < 10^6$, it can be seen that Nm will be in the range of 3×10^{13} to 3×10^{15} . This is an enormous burden for today's computing technology. For a modern computer with 1.9 GHz CPU and 256 M bytes RAM, it takes about 20 seconds to execute 10^7 times the command $n_i \leftarrow n_i + 1$ written in the MATLAB

language. It can be reasonably inferred that updating the data record for 3×10^{13} times will take about $20 \times 10^{-7} \times 3 \times 10^{13}$ seconds (i.e., about 700 days).

5.2. The Issue of Memory Space. For the conventional data structure, the total number of records is m . To execute the sample reuse algorithm or the conventional one with such data structure, each record must occupy some physical addresses. Such addresses are necessary for storing and visualizing the outcome of simulations. Of course, to obtain the outcome simulations may require a much higher amount of computer internal memory to execute the algorithm. Since m is usually a very large number, the consumption of memory to store and visualize the output of simulation can be enormous. To illustrate, consider again the example discussed in Remark 1 of Section 4. Since a floating point number occupies 2 bytes, storing a tuple of the form (k_i, n_i) needs 4 bytes. For $m \geq 3,240,000,001$, the data record will consumes $4 \times 3,240,000,001 \approx 13 \times 10^9$ bytes (i.e., about 13 giga bytes) of RAM. Such requirement, just for visualizing the outcome of the simulations, is a challenging task even for modern computers.

6. NEW TECHNIQUES OF SAMPLE REUSE

In the last section, we have shown that any algorithm using the conventional data structure suffer from the problems of the complexity of data processing and memory space. This is because, the sample size N is usually very large and the number, m , of points in the partition of uncertainty radius approaches infinity as the tolerance, ϵ , approaches zero (see Theorem 3). In this section, we shall demonstrate that, by introducing a dynamic data structure and a new sample reuse algorithm, the average requirement of memory and the computational effort devoted to data processing are absolutely bounded, independent of the tolerance, and well within the power of modern computers.

6.1. Data Structure. In order to address the memory issue and minimize the effort devoted to data processing, an appropriate data structure is critical. The key idea is to make use of the observation that, *for a set of consecutive grid points with identical records of simulation results, it suffices to store the information of the smallest and the largest grid points*. To illustrate our techniques, we enumerate, in a chronicle order of generation, the samples generated from various uncertainty bounding sets as X_1, X_2, \dots . When samples X_1, X_2, \dots, X_j have been generated, the state of the experiment is completely represented by functions $\mathbf{s}(i, j)$ and $\mathbf{v}(i, j)$, where

$$\mathbf{s}(i, j) = \sum_{k=1}^{l(i, j)} Y_i^k, \quad \mathbf{v}(i, j) = \sum_{k=1}^{l(i, j)} Z_i^k$$

with

$$Y_i^k = \begin{cases} 1 & \text{if } X_k \in \mathcal{B}_{r_i}; \\ 0 & \text{otherwise} \end{cases}$$

$$(6.1) \quad l(i, j) = \max \left\{ \ell : 1 \leq \ell \leq j, \sum_{k=1}^{\ell} Y_i^k \leq N \right\}$$

and

$$Z_i^k = \begin{cases} 1 & \text{if } X_k \in \mathcal{B}_{r_i} \text{ and } \mathbf{P} \text{ is violated for } X_k; \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \dots, m$ and $k = 1, \dots, j$. The reason we introduce variable $\mathbf{l}(i, j)$ by (6.1) is that, for grid point r_i , once N equivalent simulations are available, the subsequent simulations can be ignored. By the principle of sample reuse, $\mathbf{s}(i, j)$ and $\mathbf{v}(i, j)$ are, respectively, the accumulated numbers of samples and violations for uncertainty bounding set with radius r_i . When the experiment is completed, we have \mathbf{n} samples $X_1, X_2, \dots, X_{\mathbf{n}}$ and

$$\mathbf{s}(i, \mathbf{n}) = N, \quad \mathbb{P}(r_i) = 1 - \frac{\mathbf{v}(i, \mathbf{n})}{N}, \quad i = 1, \dots, m.$$

It can be seen that $\mathbf{s}(i, j)$ is piece-wise constant (with respect to i) and there exists a matrix S^j such that, for $i = 1, \dots, m$,

$$(6.2) \quad \mathbf{s}(i, j) = \begin{cases} [S^j]_{\ell, 2} & \text{for } [S^j]_{\ell, 1} \leq i < [S^j]_{\ell+1, 1} \quad \text{with } 1 \leq \ell \leq \kappa - 1; \\ [S^j]_{\kappa, 2} & \text{for } [S^j]_{\kappa, 1} \leq i \leq m \end{cases}$$

where κ is the number of rows of S^j and $[A]_{i,j}$ denotes the element of matrix A in the i -th row and the j -th column. Roughly speaking, *the first column of matrix S^j records the indexes of grid points for which the accumulated numbers of samples are jumping to different values. The second column of matrix S^j records the corresponding accumulated numbers of samples.*

Similarly, $\mathbf{v}(i, j)$ is piece-wise constant (with respect to i) and there exists a matrix V^j such that, for $i = 1, \dots, m$,

$$(6.3) \quad \mathbf{v}(i, j) = \begin{cases} [V^j]_{\ell, 2} & \text{for } [V^j]_{\ell, 1} \leq i < [V^j]_{\ell+1, 1} \quad \text{with } 1 \leq \ell \leq \tau - 1; \\ [V^j]_{\tau, 2} & \text{for } [V^j]_{\tau, 1} \leq i \leq m \end{cases}$$

where τ is the number of rows of V^j . Loosely speaking, *the first column of matrix V^j records the indexes of grid points for which the accumulated numbers of violations are jumping to different values. The second column of matrix V^j records the corresponding accumulated numbers of violations.*

In this paper, matrices S^j and V^j are, respectively, referred to as the *matrix of sample sizes* and the *matrix of violations*. At any stage that samples X_1, \dots, X_j have been generated, the status of the experiment is completely characterized by matrices S^j, V^j . Both matrices are of two columns but of varying number of rows in the course of experiment.

To save memory and data processing effort, we shall take advantage of the piece-wise constant property of the accumulated numbers of samples and violations. Hence, we shall construct matrices S^j and V^j when we have generated X_1, \dots, X_j . As can be seen in the sequel, such matrices can be constructed recursively. Once we have S^j and V^j , we can generate sample X_{j+1} and update S^j, V^j as S^{j+1}, V^{j+1} in accordance with equations (6.2) and (6.3).

6.2. Sample Reuse Algorithm. In this section, we present our sample reuse algorithms as follows.

Initialization: We initialize the matrices of sample sizes and violations as follows:

◇ Generate sample X_1 uniformly from uncertainty set with radius r_m .

◇ Compute j such that $X_1 \in \mathcal{B}_{r_i}$ for $j \leq i \leq m$ and $X_1 \notin \mathcal{B}_{r_i}$ for $1 \leq i \leq j-1$.

◇ Let $S^1 = [1 \quad 1]$ if $j = 1$ and $S^1 = \begin{bmatrix} 1 & 0 \\ j & 1 \end{bmatrix}$ if $j > 1$.

◇ Let $V^1 = [1 \quad 0]$ if $\mathbb{I}(X_1) = 0$ and $V^1 = S^1$ if $\mathbb{I}(X_1) = 1$, where $\mathbb{I}(X) = 1$ if the robustness requirement is violated for X and otherwise $\mathbb{I}(X) = 0$.

Sample generation: If $[S^j]_{\kappa,1} < N$ then generate sample X_{j+1} uniformly from uncertainty set with radius r_m , otherwise generate sample X_{j+1} uniformly from uncertainty set with radius $[S^j]_{\kappa-1,1}$.

Updating matrices: Update S^j as S^{j+1} by the method described in Section 6.2.1. If $\mathbb{I}(X_{j+1}) = 0$ then let $V^{j+1} = V^j$, otherwise update V^j as V^{j+1} by the method described in Section 6.2.2.

Stopping criterion: The sampling process is terminated if S^j has only one row and $[S^j]_{1,2} = N$.

6.2.1. *Sample Sizes Tracking.* In this section, we describe how to update the matrix of sample sizes. The key idea is to ensure condition (6.2). Let κ be the number of rows of S^j . We proceed as follows.

Step (1): Compute an index j^* such that $X_{j+1} \in \mathcal{B}_{r_i}$ for $j^* \leq i \leq m$ and $X_{j+1} \notin \mathcal{B}_{r_i}$ for $1 \leq i \leq j^* - 1$ (Note that explicit formulas for computing j^* are available when using uniform or geometric grid scheme).

Step (2): Modify S^j as a temporary matrix \widehat{S}^{j+1} based on the following three cases.

Case (1): $[S^j]_{\ell^*,1} < j^* < [S^j]_{\ell^*+1,1}$ for some $\ell^* \in \{1, \dots, \kappa-1\}$;

Case (2): $j^* = [S^j]_{\ell^*,1}$ for some $\ell^* \in \{1, \dots, \kappa\}$;

Case (3): $j^* > [S^j]_{\kappa,1}$.

In Case (1), define \widehat{S}^{j+1} as a $(\kappa+1) \times 2$ matrix such that

$$\begin{aligned} [\widehat{S}^{j+1}]_{\ell,1} &= [S^j]_{\ell,1}, & [\widehat{S}^{j+1}]_{\ell,2} &= [S^j]_{\ell,2}, & \ell &= 1, \dots, \ell^* \\ [\widehat{S}^{j+1}]_{\ell^*+1,1} &= j^*, & [\widehat{S}^{j+1}]_{\ell^*+1,2} &= 1 + [S^j]_{\ell^*,2} \\ [\widehat{S}^{j+1}]_{\ell+1,1} &= [S^j]_{\ell,1}, & [\widehat{S}^{j+1}]_{\ell+1,2} &= 1 + [S^j]_{\ell,2}, & \ell &= \ell^* + 1, \dots, \kappa. \end{aligned}$$

In Case (2), define \widehat{S}^{j+1} as a $\kappa \times 2$ matrix such that

$$\begin{aligned} [\widehat{S}^{j+1}]_{\ell,1} &= [S^j]_{\ell,1}, & [\widehat{S}^{j+1}]_{\ell,2} &= [S^j]_{\ell,2}, & \ell &= 1, \dots, \ell^* - 1 \\ [\widehat{S}^{j+1}]_{\ell^*,1} &= [S^j]_{\ell^*,1}, & [\widehat{S}^{j+1}]_{\ell^*,2} &= 1 + [S^j]_{\ell^*,2}, & \ell &= \ell^*, \dots, \kappa. \end{aligned}$$

In Case (3), define \widehat{S}^{j+1} as a $(\kappa+1) \times 2$ matrix such that

$$\begin{aligned} [\widehat{S}^{j+1}]_{\ell,1} &= [S^j]_{\ell,1}, & [\widehat{S}^{j+1}]_{\ell,2} &= [S^j]_{\ell,2}, & \ell &= 1, \dots, \kappa \\ [\widehat{S}^{j+1}]_{\kappa+1,1} &= j^*, & [\widehat{S}^{j+1}]_{\kappa+1,2} &= 1 + [S^j]_{\kappa,2}. \end{aligned}$$

Step (3): Let $\hat{\kappa}$ denote the number of rows of \hat{S}^{j+1} . If $[\hat{S}^{j+1}]_{\hat{\kappa},2} < N$ then let $S^{j+1} = \hat{S}^{j+1}$, otherwise find index ℓ_* by a bisection search such that $[\hat{S}^j]_{\ell_*,2} < N \leq [\hat{S}^j]_{\ell_*,2}$ and define S^{j+1} as an $\ell_* \times 2$ matrix such that

$$\begin{aligned} [S^{j+1}]_{\ell,1} &= [\hat{S}^{j+1}]_{\ell,1}, & [S^{j+1}]_{\ell,2} &= [\hat{S}^{j+1}]_{\ell,2}, & \ell &= 1, \dots, \ell_* - 1 \\ [S^{j+1}]_{\ell_*,1} &= [\hat{S}^{j+1}]_{\ell_*,1}, & [S^{j+1}]_{\ell_*,2} &= [\hat{S}^{j+1}]_{\ell_*,2}. \end{aligned}$$

6.2.2. Violations Tracking. In this section, we describe how to update the matrix of violations in the case of $\mathbb{I}(X_{j+1}) = 1$. The key idea is to ensure condition (6.3). Let κ be the number of rows of S^j . Let τ be the number of rows of V^j . Let j^* be the index obtained in the process of updating S^j such that $X_{j+1} \in \mathcal{B}_{r_i}$ for $j^* \leq i \leq m$ and $X_{j+1} \notin \mathcal{B}_{r_i}$ for $1 \leq i \leq j^* - 1$. We proceed as follows.

Step (i): Identify the *maximal* index ι such that the experiment for uncertainty radius r_ι has not been completed by the following method.

◇ If $[S^j]_{\kappa,2} < N$, then let $\iota = \kappa$, otherwise find ι by a bisection search such that $[V^j]_{\iota,1} < [S^j]_{\kappa,1}$, $[V^j]_{\iota+1,1} \geq [S^j]_{\kappa,1}$.

Step (ii): Modify V^j as a temporary matrix \hat{V}^j based on the following two cases.

Case(a) : $[S^j]_{\kappa,2} < N$ or $[S^j]_{\kappa,2} = N$, $[V^j]_{\iota+1,1} = [S^j]_{\kappa,1}$.
Case(b) : $[S^j]_{\kappa,2} \geq N$ and the index ι guarantees $[V^j]_{\iota+1,1} > [S^j]_{\kappa,1}$.

In Case (a), we define $\hat{V}^j = V^j$. In Case (b), we define \hat{V}^j as a $(\tau+1) \times 2$ matrix such that

$$\begin{aligned} [\hat{V}^j]_{\ell,1} &= [V^j]_{\ell,1}, & [\hat{V}^j]_{\ell,2} &= [V^j]_{\ell,2}, & \ell &= 1, \dots, \iota \\ [\hat{V}^j]_{\iota+1,1} &= [S^j]_{\kappa,1}, & [\hat{V}^j]_{\iota+1,2} &= [V^j]_{\iota,2} \\ [\hat{V}^j]_{\ell+1,1} &= [V^j]_{\ell,1}, & [\hat{V}^j]_{\ell+1,2} &= [V^j]_{\ell,2}, & \ell &= \iota+1, \dots, \tau. \end{aligned}$$

Step (iii): Obtain V^{j+1} by modifying \hat{V}^j based on the following three cases.

Case (i): $[\hat{V}^j]_{\ell^*,1} < j^* < [\hat{V}^j]_{\ell^*+1,1}$ for some $\ell^* \in \{1, \dots, \iota-1\}$;
Case (ii): $j^* = [\hat{V}^j]_{\ell^*,1}$ for some $\ell^* \in \{1, \dots, \iota\}$;
Case (iii): $j^* > [\hat{V}^j]_{\iota,1}$.

Let $\hat{\tau}$ be the number of rows of \hat{V}^j . In Case (i), define V^{j+1} as a $(\hat{\tau}+1) \times 2$ matrix such that

$$\begin{aligned} [V^{j+1}]_{\ell,1} &= [\hat{V}^j]_{\ell,1}, & [V^{j+1}]_{\ell,2} &= [\hat{V}^j]_{\ell,2}, & \ell &= 1, \dots, \ell^* \\ [V^{j+1}]_{\ell^*+1,1} &= j^*, & [V^{j+1}]_{\ell^*+1,2} &= 1 + [\hat{V}^j]_{\ell^*,2} \\ [V^{j+1}]_{\ell+1,1} &= [\hat{V}^j]_{\ell,1}, & [V^{j+1}]_{\ell+1,2} &= 1 + [\hat{V}^j]_{\ell,2}, & \ell &= \ell^*+1, \dots, \iota \\ [V^{j+1}]_{\ell+1,1} &= [\hat{V}^j]_{\ell,1}, & [V^{j+1}]_{\ell+1,2} &= [\hat{V}^j]_{\ell,2}, & \ell &= \iota+1, \dots, \hat{\tau}. \end{aligned}$$

In Case (ii), define V^{j+1} as a $\hat{\tau} \times 2$ matrix such that

$$\begin{aligned} [V^{j+1}]_{\ell,1} &= [\hat{V}^j]_{\ell,1}, & [V^{j+1}]_{\ell,2} &= [\hat{V}^j]_{\ell,2}, & \ell &= 1, \dots, \ell^* - 1 \\ [V^{j+1}]_{\ell,1} &= [\hat{V}^j]_{\ell,1}, & [V^{j+1}]_{\ell,2} &= 1 + [\hat{V}^j]_{\ell,2}, & \ell &= \ell^*, \dots, \iota \\ [V^{j+1}]_{\ell,1} &= [\hat{V}^j]_{\ell,1}, & [V^{j+1}]_{\ell,2} &= [\hat{V}^j]_{\ell,2}, & \ell &= \iota+1, \dots, \hat{\tau}. \end{aligned}$$

In Case (iii), define V^{j+1} as a $(\hat{\tau} + 1) \times 2$ matrix such that

$$\begin{aligned} [V^{j+1}]_{\ell,1} &= [\hat{V}^j]_{\ell,1}, & [V^{j+1}]_{\ell,2} &= [\hat{V}^j]_{\ell,2}, & \ell &= 1, \dots, \iota \\ [V^{j+1}]_{\iota+1,1} &= j^*, & [V^{j+1}]_{\iota+1,2} &= 1 + [\hat{V}^j]_{\iota,2} \\ [V^{j+1}]_{\ell+1,1} &= [\hat{V}^j]_{\ell,1}, & [V^{j+1}]_{\ell+1,2} &= [\hat{V}^j]_{\ell,2}, & \ell &= \iota + 1, \dots, \hat{\tau}. \end{aligned}$$

6.3. Complexity of Data Processing and Memory. It can be seen that the memory requirement and the computation due to data processing are determined by the sizes of matrices S^j and V^j . To quantify the complexity, we have the following results.

Theorem 5. *For any j , the following statements hold true:*

- (I): *The number of rows of matrix S^j is no more than N ;*
- (II): *The expected number of rows of matrix V^j is no greater than*

$$(6.4) \quad 1 + N \left[P_e(a) + 2d \int_{\frac{a}{\hbar}}^a \frac{P_e(x)}{x} dx \right] \leq 1 + NP_e(a) (1 + 2d \ln \hbar)$$

where $P_e(x) = 1 - \min_{y \in [\frac{a}{\lambda}, x]} \mathbb{P}(y)$, $\forall x \in [\frac{a}{\lambda}, a]$ and

$$\hbar = \max \left(\min \left(\lambda, \frac{a}{\rho_0} \right), 1 \right)$$

with $\rho_0 = \sup\{r \mid \mathbb{P}(r) = 1\}$.

See Appendix F for a proof. We now revisit the robustness analysis problem discussed in Remark 1 of Section 4 from the perspective of memory complexity. Assume that each data record (k_i, n_i) (or each row of V^j) occupies 4 bytes of computer internal memory (RAM). As illustrated in Section 5.2, when using the conventional data structure, it takes 13G (giga bytes) of RAM to save the data and visualize the results. On the other hand, in our new algorithm, if the smallest proportion is $p_* = \min_{r \in [\frac{a}{\lambda}, a]} \mathbb{P}(r) > 0.999$ and $\hbar < \frac{3}{2}$, the RAM requirement will be equivalent to

$$\begin{aligned} & 4 \times [1 + (1 - p_*) (1 + 2d \ln \hbar) N] \\ &= 4 \times \left[1 + (1 - 0.999) \times \left(1 + 2 \times 1800 \times \ln \frac{3}{2} \right) \times 10^6 \right] \\ &< 6.2 \times 10^6 \text{ bytes} \approx 6.2 \text{ M bytes}. \end{aligned}$$

It can be seen that such requirement of memory is extremely low as compared to that of the conventional method. Theorem 5 also reveals that the complexity of data processing is very low.

6.4. Confidence Band. To be useful, any numerical techniques should provide a method for error assessment. Monte Carlo simulation is no exception. The following results allows us to construct confidence band for the robustness curve. Such post-experimental statistical inference can remedy the conservatism of *a priori* choice of sample size N based on the Chernoff bound. In order to overcome the computational complexity of the Clopper-Pearson's confidence interval [9], we have developed new methods to facilitate the construction of the confidence band.

Theorem 6. Let $\delta \in (0, 1)$. Let $\mathcal{L}(k) = \frac{k}{N} + \frac{3}{4} \frac{1 - \frac{2k}{N} - \sqrt{1 + 4\theta \frac{k(1 - \frac{k}{N})}{1 + \theta N}}}{1 + \theta N}$ and $\mathcal{U}(k) = \frac{k}{N} + \frac{3}{4} \frac{1 - \frac{2k}{N} + \sqrt{1 + 4\theta \frac{k(1 - \frac{k}{N})}{1 + \theta N}}}{1 + \theta N}$ with $\theta = \frac{9}{8 \ln \frac{2}{\delta}}$. Let $\zeta = \frac{r - r_i}{r_{i+1} - r_i}$. Let

$$K_i = N - \mathbf{v}(i, \mathbf{n}), \quad i = 1, \dots, m.$$

Let $\varsigma = 1 - \frac{g(r_*)}{r_{i+1} - r_i}$. Define $\overline{\mathbb{P}}(r) = \zeta \mathcal{U}(K_i) + (1 - \zeta) \mathcal{U}(K_{i+1}) + \varsigma$ and $\underline{\mathbb{P}}(r) = \zeta \mathcal{L}(K_i) + (1 - \zeta) \mathcal{L}(K_{i+1}) - \varsigma$. Then

$$\Pr\{\underline{\mathbb{P}}(r) < \mathbb{P}(r) < \overline{\mathbb{P}}(r), \forall r \in [r_i, r_{i+1}]\} > 1 - \delta.$$

See Appendix G for a proof. The family of intervals $[\underline{\mathbb{P}}(r), \overline{\mathbb{P}}(r)]$, $r \in [a/\lambda, a]$ is referred to as the *confidence band*. It is important to note that the confidence band can be efficiently constructed by making use of the piece-wise constant property of $\mathbf{v}(i, \mathbf{n})$. It can be shown that the computational complexity of constructing the confidence band is also absolutely bounded.

7. CONCLUSION

It is possible to make a case for the statement that the probabilistic robustness analysis is essentially the study of the robustness function, especially about its probabilistic implications, efficient evaluation and computational complexity. We have addressed these issues in this paper. In particular, we have developed randomized algorithms which offer more insights for system robustness. We rigorously show that, in both aspects of computer running time and memory requirement, the complexity of such randomized algorithms is not only linear in the dimension of uncertainty space, but also surprisingly low. While the complexity of conventional method grows linearly with the number of grid points and the error due to interpolation is not well controlled, our techniques completely resolve such issues. In short, our method guarantees accuracy and efficiency.

APPENDIX A. PROOF OF THEOREM 1

We first establish a basic inequality that will be used to prove the theorem.

Lemma 1. For any $x > 1$,

$$\frac{1}{x} + \ln x > 1.$$

Proof. Let

$$f(x) = \frac{1}{x} + \ln x.$$

Then $f(1) = 1$ and

$$\frac{d f(x)}{dx} = \frac{x - 1}{x^2} > 0, \quad \forall x > 1.$$

It follows that $f(x) > 1$, $\forall x > 1$. □

Now we are in the position to prove Theorem 1. Observing that

$$\left(\frac{r_m}{r_1}\right)^d = \prod_{i=1}^{m-1} \left(\frac{r_{i+1}}{r_i}\right)^d,$$

we have

$$\ln \left(\frac{r_m}{r_1}\right)^d = \sum_{i=1}^{m-1} \ln \left(\frac{r_{i+1}}{r_i}\right)^d.$$

Therefore,

$$\sum_{i=1}^{m-1} \left(\frac{r_i}{r_{i+1}} \right)^d + \ln \left(\frac{r_m}{r_1} \right)^d = \sum_{i=1}^{m-1} \left[\frac{1}{\left(\frac{r_{i+1}}{r_i} \right)^d} + \ln \left(\frac{r_{i+1}}{r_i} \right)^d \right].$$

Since $\left(\frac{r_{i+1}}{r_i} \right)^d > 1$, $i = 1, \dots, m-1$, it follows from Lemma 1 that

$$\frac{1}{\left(\frac{r_{i+1}}{r_i} \right)^d} + \ln \left(\frac{r_{i+1}}{r_i} \right)^d > 1, \quad i = 1, \dots, m-1.$$

Hence,

$$\sum_{i=1}^{m-1} \left(\frac{r_i}{r_{i+1}} \right)^d + \ln \left(\frac{r_m}{r_1} \right)^d > m-1,$$

or equivalently,

$$m - \sum_{i=1}^{m-1} \left(\frac{r_i}{r_{i+1}} \right)^d < 1 + \ln \left(\frac{r_m}{r_1} \right)^d = 1 + d \ln \lambda.$$

Finally, by Theorem 1 of [6] and the definition of m_{eq} , we have

$$m_{\text{eq}} = m - \sum_{i=1}^{m-1} \left(\frac{r_i}{r_{i+1}} \right)^d < 1 + d \ln \lambda.$$

APPENDIX B. PROOF OF THEOREM 2

To prove the theorem, we need some preliminary results. It is derived in [2] that $\left| \frac{d\mathbb{P}(r)}{dr} \right| < \frac{2d}{r}$ when $\mathbb{P}(\cdot)$ is differentiable. The following lemma indicates that the bound on the rate of variation of $\mathbb{P}(\cdot)$ can be much tighter.

Lemma 2. *For arbitrary robustness requirement,*

$$|\mathbb{P}(r + \Delta r) - \mathbb{P}(r)| \leq 1 - \left(1 + \frac{\Delta r}{r} \right)^{-d} < \frac{d}{r} \Delta r$$

for any $r > 0$ and any $\Delta r > 0$.

Proof. Let $\mathcal{Q}_r \subseteq \mathcal{B}_r$ be the set such that the robustness requirement is satisfied. Let

$$I_1 = \frac{\text{vol}(\mathcal{Q}_{r+\Delta r})}{\text{vol}(\mathcal{B}_{r+\Delta r})} - \frac{\text{vol}(\mathcal{Q}_r)}{\text{vol}(\mathcal{B}_{r+\Delta r})}, \quad I_2 = \frac{\text{vol}(\mathcal{Q}_r)}{\text{vol}(\mathcal{B}_{r+\Delta r})} - \frac{\text{vol}(\mathcal{Q}_r)}{\text{vol}(\mathcal{B}_r)}.$$

Let “ \setminus ” denote the operation of set minus. Observing that $\mathcal{Q}_{r+\Delta r} \setminus \mathcal{Q}_r \subseteq \mathcal{B}_{r+\Delta r} \setminus \mathcal{B}_r$, we have $\text{vol}(\mathcal{Q}_{r+\Delta r}) - \text{vol}(\mathcal{Q}_r) \leq \text{vol}(\mathcal{B}_{r+\Delta r}) - \text{vol}(\mathcal{B}_r)$. Using this fact and the identity $\text{vol}(\mathcal{B}_r) = r^d \text{vol}(\mathcal{B}_1)$, we have

$$0 \leq I_1 \leq \frac{\text{vol}(\mathcal{B}_{r+\Delta r}) - \text{vol}(\mathcal{B}_r)}{\text{vol}(\mathcal{B}_{r+\Delta r})} = 1 - \left(1 + \frac{\Delta r}{r} \right)^{-d}$$

and

$$- \left[1 - \left(1 + \frac{\Delta r}{r} \right)^{-d} \right] \leq - \frac{\text{vol}(\mathcal{Q}_r)}{\text{vol}(\mathcal{B}_r)} \frac{\text{vol}(\mathcal{B}_{r+\Delta r}) - \text{vol}(\mathcal{B}_r)}{\text{vol}(\mathcal{B}_{r+\Delta r})} = I_2 \leq 0.$$

Therefore, $|\mathbb{P}(r + \Delta r) - \mathbb{P}(r)| = |I_1 + I_2| \leq 1 - \left(1 + \frac{\Delta r}{r}\right)^{-d} < \frac{d}{r} \Delta r$ where the last inequality follows from inequality $1 - \left(1 + \frac{x}{r}\right)^{-d} < \frac{dx}{r}$, $\forall x > 0$. To prove this inequality, we can define function $h(x) \stackrel{\text{def}}{=} 1 - \left(1 + \frac{x}{r}\right)^{-d} - \frac{dx}{r}$, $x > 0$ and check that $h(0) = 0$ and $\frac{\partial h(x)}{\partial x} = \frac{d}{r} \left[\left(1 + \frac{x}{r}\right)^{-(d+1)} - 1 \right] < 0$, $\forall x > 0$. \square

We are now in the position to prove the theorem. It can be shown that

$$\begin{aligned} |\mathbb{P}(r) - \mathbb{P}^*(r)| &= \left| \frac{(r_{i+1} - r)[\mathbb{P}(r) - \mathbb{P}(r_i)] + (r - r_i)[\mathbb{P}(r) - \mathbb{P}(r_{i+1})]}{r_{i+1} - r_i} \right| \\ (B.1) \quad &\leq \frac{(r_{i+1} - r)|\mathbb{P}(r) - \mathbb{P}(r_i)| + (r - r_i)|\mathbb{P}(r) - \mathbb{P}(r_{i+1})|}{r_{i+1} - r_i}. \end{aligned}$$

By Lemma 2 and inequality (B.1), we have

$$\begin{aligned} |\mathbb{P}(r) - \mathbb{P}^*(r)| &\leq \frac{(r_{i+1} - r) \left[1 - \left(\frac{r}{r_i}\right)^{-d} \right] + (r - r_i) \left[1 - \left(\frac{r_{i+1}}{r}\right)^{-d} \right]}{r_{i+1} - r_i} \\ &= 1 - \frac{g(r)}{r_{i+1} - r_i}. \end{aligned}$$

Note that

$$\frac{\partial g(r)}{\partial r} = \Phi(r) - \Psi(r)$$

where

$$\Phi(r) = \left(\frac{r_{i+1}}{r}\right)^{-d} \left[1 + \left(1 - \frac{r_i}{r}\right) d \right], \quad \Psi(r) = \left(\frac{r}{r_i}\right)^{-d} \left[1 + \left(\frac{r_{i+1}}{r} - 1\right) d \right].$$

It can be verified that

$$\begin{aligned} \Phi(r_i) &= \left(\frac{r_{i+1}}{r_i}\right)^{-d} < 1, \quad \Psi(r_i) = 1 + \left(\frac{r_{i+1}}{r_i} - 1\right) d > 1, \quad \frac{\partial g(r)}{\partial r} \Big|_{r=r_i} < 0, \\ \Phi(r_{i+1}) &= 1 + \left(1 - \frac{r_i}{r_{i+1}}\right) d > 1, \quad \Psi(r_{i+1}) = \left(\frac{r_{i+1}}{r_i}\right)^{-d} < 1, \quad \frac{\partial g(r)}{\partial r} \Big|_{r=r_{i+1}} > 0. \end{aligned}$$

It can be checked that $\Phi(r)$ is a monotone increasing function of r and that $\Psi(r)$ is a monotone decreasing function of r . Hence, $\frac{\partial g(r)}{\partial r}$ is a monotone increasing function of r . Moreover, there exists a unique $r_\star \in (r_i, r_{i+1})$ such that $\frac{\partial g(r)}{\partial r} \Big|_{r=r_\star} = 0$, i.e., $\Phi(r_\star) = \Psi(r_\star)$. Furthermore, $g(r)$ is a convex function of r . Consequently,

$$\min_{r \in [r_i, r_{i+1}]} g(r) = g(r_\star)$$

and we have shown

$$|\mathbb{P}(r) - \mathbb{P}^*(r)| \leq 1 - \frac{g(r_\star)}{r_{i+1} - r_i} \quad \forall r \in [r_i, r_{i+1}].$$

Since $\frac{\partial g(r)}{\partial r}$ is a monotone increasing function of r , we can compute r_\star by a bisection search over interval (r_i, r_{i+1}) .

By Lemma 2 and inequality (B.1), we have

$$\begin{aligned}
|\mathbb{P}(r) - \mathbb{P}^*(r)| &\leq \frac{(r_{i+1} - r)(r - r_i) \frac{d}{r_i} + (r - r_i)(r_{i+1} - r) \frac{d}{r}}{r_{i+1} - r_i} \\
&\leq \frac{(r_{i+1} - r)(r - r_i) \frac{d}{r_i} + (r - r_i)(r_{i+1} - r) \frac{d}{r_i}}{r_{i+1} - r_i} \\
&\leq \frac{2d}{r_i(r_{i+1} - r_i)} \max_{r \in [r_i, r_{i+1}]} (r_{i+1} - r)(r - r_i) \\
&= \frac{2d}{r_i(r_{i+1} - r_i)} \frac{(r_{i+1} - r_i)^2}{4} \\
&= \frac{d(r_{i+1} - r_i)}{2r_i}.
\end{aligned}$$

APPENDIX C. PROOF OF THEOREM 3

By Theorem 2, $|\mathbb{P}(r) - \mathbb{P}^*(r)| \leq \frac{d(r_{i+1} - r_i)}{2r_i}$, $\forall r \in [r_i, r_{i+1}]$. Thus, it suffices to show $\frac{d(r_{i+1} - r_i)}{2r_i} < \epsilon$, i.e.,

$$(C.1) \quad \frac{r_{i+1}}{r_i} < 1 + \frac{2\epsilon}{d}.$$

By definition (4.1), for $i = 1, \dots, m-1$,

$$\begin{aligned}
\frac{r_{i+1}}{r_i} &= \frac{a - \frac{(m-i-1)(\lambda-1)}{(m-1)\lambda}a}{a - \frac{(m-i)(\lambda-1)}{(m-1)\lambda}a} \\
&= 1 + \frac{\lambda-1}{m-1 + (\lambda-1)(i-1)} \\
&\leq 1 + \frac{\lambda-1}{m-1}.
\end{aligned}$$

By virtue of (C.1), to guarantee that the gridding error is less than ϵ , it suffices to ensure $1 + \frac{\lambda-1}{m-1} < 1 + \frac{2\epsilon}{d}$, i.e., $m > 1 + \frac{d(\lambda-1)}{2\epsilon}$. Hence, it suffices to have

$$m \geq 2 + \left\lfloor \frac{(\lambda-1)d}{2\epsilon} \right\rfloor.$$

It can be verified that

$$\frac{r_i}{r_{i+1}} = 1 - \frac{1}{\frac{m-1}{\lambda-1} + i}, \quad i = 1, \dots, m-1.$$

By Theorem 1 of [6], the sample reuse factor is given by

$$\begin{aligned}
\mathcal{F}_{\text{reuse}} &= \frac{m}{m - \sum_{i=1}^{m-1} \left(\frac{r_i}{r_{i+1}} \right)^d} \\
&= \frac{m}{m - \sum_{i=1}^{m-1} \left(1 - \frac{1}{\frac{m-1}{\lambda-1} + i} \right)^d}.
\end{aligned}$$

Therefore,

$$m_{\text{eq}}(\epsilon) = \frac{m}{\mathcal{F}_{\text{reuse}}} = m - \sum_{i=1}^{m-1} \left(1 - \frac{1}{\frac{m-1}{\lambda-1} + i} \right)^d.$$

APPENDIX D. PROOF OF THEOREM 4

By virtue of (4.2), we have $\frac{r_{i+1}}{r_i} = \lambda^{\frac{1}{m-1}}$. Hence, by (C.1), it suffices to show $\lambda^{\frac{1}{m-1}} < 1 + \frac{2\epsilon}{d}$, which can be reduced to $m > 1 + \frac{\ln \lambda}{\ln(1 + \frac{2\epsilon}{d})}$. This inequality is equivalent to $m \geq 2 + \left\lfloor \frac{\ln \lambda}{\ln(1 + \frac{2\epsilon}{d})} \right\rfloor$. By equation (3.1) and Theorem 1 of [6], we have $\mathbb{E}[\mathbf{n}] = \left[m - (m-1) \left(\frac{1}{\lambda} \right)^{\frac{d}{m-1}} \right] N$ and hence obtain $m_{\text{eq}}(\epsilon)$. Note that

$$\mathcal{F}_{\text{reuse}} = \frac{m}{m_{\text{eq}}(\epsilon)} > \frac{m}{1 + d \ln \lambda} = \frac{2 + \left\lfloor \frac{\ln \lambda}{\ln(1 + \frac{2\epsilon}{d})} \right\rfloor}{1 + d \ln \lambda} > \frac{\frac{\ln \lambda}{\ln(1 + \frac{2\epsilon}{d})}}{1 + d \ln \lambda}.$$

Making use of the inequality $\ln(1+x) < x$, $\forall x > 0$, we have $\ln(1 + \frac{2\epsilon}{d}) < \frac{2\epsilon}{d}$. Therefore,

$$\mathcal{F}_{\text{reuse}} > \frac{\frac{\ln \lambda}{\frac{2\epsilon}{d}}}{1 + d \ln \lambda} = \frac{1}{2\epsilon} \left(1 - \frac{1}{1 + d \ln \lambda} \right).$$

APPENDIX E. PROOF OF THEOREM 5

Proof of statement (I): Obviously, $[S^j]_{1,2} \geq 1$, $[S^j]_{\kappa,2} \leq N$. From the rules of sampling, we can perform induction with respect to j and have $[S^j]_{\ell+1,2} - [S^j]_{\ell,2} \geq 1$, $\ell = 1, \dots, \kappa - 1$. Observing that

$$\begin{aligned} [S^j]_{\kappa,2} &= [S^j]_{1,2} + \sum_{\ell=1}^{\kappa-1} ([S^j]_{\ell+1,2} - [S^j]_{\ell,2}) \\ &\geq 1 + \sum_{\ell=1}^{\kappa-1} ([S^j]_{\ell+1,2} - [S^j]_{\ell,2}) \\ &\geq 1 + \kappa - 1 \\ &= \kappa, \end{aligned}$$

we have $\kappa \leq [S^j]_{\kappa,2} \leq N$.

Proof of statement (II): We need some preliminary results.

Lemma 3. Let $1 \leq i \leq m-1$. Then

$$\left| \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - \frac{d(r_{i+1} - r_i)}{r_{i+1}} \right| \leq \frac{d(d-1)}{2} \left(\frac{r_{i+1} - r_i}{r_{i+1}} \right)^2.$$

Proof. Note that

$$\begin{aligned} \left| \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - \frac{d(r_{i+1} - r_i)}{r_{i+1}} \right| &= \left| \left(\frac{r_i}{r_{i+1}} \right)^d - \left(1 - \frac{d(r_{i+1} - r_i)}{r_{i+1}} \right) \right| \\ &= \left| (1-t)^d - (1-dt) \right| \end{aligned}$$

where $t = \frac{r_{i+1} - r_i}{r_{i+1}}$. It can be checked that $\left| (1-t)^d - (1-dt) \right| = \frac{d(d-1)}{2} t^2$ for $d = 1, 2$. For $d \geq 3$, by Taylor's expansion formula, there exists $\xi \in (0, t)$ such that

$$(1-t)^d = 1 - dt + \frac{d(d-1)}{2} (1-\xi)^{d-2} t^2.$$

Observing that $0 < t < 1$ since $0 < r_i < r_{i+1}$, we hence have $0 < (1 - \xi)^{d-2} < 1$ and $\left| (1-t)^d - (1-dt) \right| < \frac{d(d-1)}{2} t^2$ for $d \geq 3$. Therefore, for any $d \geq 1$,

$$\left| \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - \frac{d(r_{i+1} - r_i)}{r_{i+1}} \right| \leq \frac{d(d-1)}{2} t^2 = \frac{d(d-1)}{2} \left(\frac{r_{i+1} - r_i}{r_{i+1}} \right)^2.$$

□

Lemma 4. *Define the maximum gap between grid points as*

$$\varpi = \max_{1 \leq i \leq m-1} (r_{i+1} - r_i).$$

Then

$$\sum_{i=1}^{m-1} \left(\frac{r_{i+1} - r_i}{r_{i+1}} \right)^2 < \frac{\lambda(\lambda-1)\varpi}{a}.$$

Proof. Note that

$$\sum_{i=1}^{m-1} \left(\frac{r_{i+1} - r_i}{r_{i+1}} \right)^2 \leq \varpi \sum_{i=1}^{m-1} \frac{r_{i+1} - r_i}{r_{i+1}^2} < \varpi \sum_{i=1}^{m-1} \frac{r_{i+1} - r_i}{\left(\frac{a}{\lambda}\right)^2}.$$

By successive cancelation,

$$\sum_{i=1}^{m-1} (r_{i+1} - r_i) = r_m - r_1 = a - \frac{a}{\lambda}.$$

Hence,

$$\sum_{i=1}^{m-1} \left(\frac{r_{i+1} - r_i}{r_{i+1}} \right)^2 < \varpi \sum_{i=1}^{m-1} \frac{r_{i+1} - r_i}{\left(\frac{a}{\lambda}\right)^2} = \varpi \frac{a - \frac{a}{\lambda}}{\left(\frac{a}{\lambda}\right)^2} = \frac{\lambda(\lambda-1)\varpi}{a}.$$

□

Lemma 5. *The expected number of rows of the matrix of violations $V^{\mathbf{n}}$ is no greater than $1 + NP_e(a) + 2N \sum_{j=1}^{m-1} P_e(r_j) \left[1 - \left(\frac{r_j}{r_{j+1}} \right)^d \right]$.*

Proof. Let $X_1^j, \dots, X_{\mathbf{n}_j}^j$ be the samples generated from uncertainty set with radius r_j . Let $Y_i^j = \mathbb{I}(X_i^j)$, $i = 1, \dots, \mathbf{n}_j$. By the principle of sample reuse, the value of \mathbf{n}_j depends only on the samples generated from uncertainty sets with radius r_k , $j+1 \leq k \leq m$. Consequently, event $\{\mathbf{n}_j = \nu\}$ is independent of event $\{Y_i^j = 1\}$ and $\Pr\{Y_i^j = 1, \mathbf{n}_j = \nu\} = \Pr\{Y_i^j = 1\} \Pr\{\mathbf{n}_j = \nu\}$. By the definitions of Y_i^j and $P_e(\cdot)$, we have

$\Pr\{Y_i^j = 1\} = 1 - \mathbb{P}(r_j) \leq P_e(r_j)$. Therefore,

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^{\mathbf{n}_j} Y_i^j \right] &= \sum_{\nu=1}^N \sum_{i=1}^{\nu} \Pr\{Y_i^j = 1, \mathbf{n}_j = \nu\} \\ &= \sum_{\nu=1}^N \sum_{i=1}^{\nu} \Pr\{Y_i^j = 1\} \Pr\{\mathbf{n}_j = \nu\} \\ &\leq \sum_{\nu=1}^N \nu P_e(r_j) \Pr\{\mathbf{n}_j = \nu\} \\ &= P_e(r_j) \mathbb{E}[\mathbf{n}_j] \end{aligned}$$

for $j = 1, \dots, m$. We now consider $V^{\mathbf{n}}$ with $\mathbf{n} = \sum_{i=1}^m \mathbf{n}_i$. By the mechanism of the sample reuse algorithms, for $j = 1, \dots, m-1$, every new sample from uncertainty set with radius r_j at most creates $2Y_i^j$, $i = 1, \dots, \mathbf{n}_j$ new rows for the matrix of violations (see Section 6.2.2). Note that X_1 create at most $1 + Y_1^m$ rows for V^1 . Every new sample from uncertainty set with radius r_m at most creates Y_i^m , $i = 2, \dots, \mathbf{n}_m$ new rows for the matrix of violations. Hence

$$\begin{aligned} &\mathbb{E}[\text{The number of rows of matrix } V^{\mathbf{n}}] \\ &\leq 1 + \mathbb{E} \left[\sum_{i=1}^{\mathbf{n}_m} Y_i^m \right] + 2 \mathbb{E} \left[\sum_{j=1}^{m-1} \sum_{i=1}^{\mathbf{n}_j} Y_i^j \right] \\ &\leq 1 + P_e(r_m) \mathbb{E}[\mathbf{n}_m] + 2 \sum_{j=1}^{m-1} P_e(r_j) \mathbb{E}[\mathbf{n}_j]. \end{aligned}$$

By Lemma 6 of [6], we have

$$(E.1) \quad \mathbb{E}[\mathbf{n}_j] = \left[1 - \left(\frac{r_j}{r_{j+1}} \right)^d \right] N, \quad j = 1, \dots, m-1.$$

By (E.1) and using the fact that $\mathbb{E}[\mathbf{n}_m] = N$, $P_e(r_m) = P_e(a)$, we have

$$\mathbb{E}[\text{The number of rows of matrix } V^{\mathbf{n}}] \leq 1 + NP_e(a) + 2N \sum_{j=1}^{m-1} P_e(r_j) \left[1 - \left(\frac{r_j}{r_{j+1}} \right)^d \right].$$

□

Lemma 6. *For any grid scheme,*

$$\begin{aligned} &\left| \sum_{i=1}^{m-1} P_e(r_i) \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - d \sum_{i=1}^{m-1} \frac{P_e(r_{i+1})(r_{i+1} - r_i)}{r_{i+1}} \right| \\ &< \frac{d(d-1)\lambda(\lambda-1)\varpi}{2a} + \frac{\lambda d}{a} \sum_{i=1}^{m-1} P_e(r_{i+1})(r_{i+1} - r_i) - \frac{\lambda d}{a} \sum_{i=1}^{m-1} P_e(r_i)(r_{i+1} - r_i). \end{aligned}$$

Proof. Note that

$$\begin{aligned}
& \left| \sum_{i=1}^{m-1} P_e(r_i) \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - d \sum_{i=1}^{m-1} \frac{P_e(r_{i+1})(r_{i+1} - r_i)}{r_{i+1}} \right| \\
& \leq \sum_{i=1}^{m-1} \left| P_e(r_i) \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - d \frac{P_e(r_i)(r_{i+1} - r_i)}{r_{i+1}} \right| \\
& \quad + \sum_{i=1}^{m-1} \left| d \frac{P_e(r_i)(r_{i+1} - r_i)}{r_{i+1}} - d \frac{P_e(r_{i+1})(r_{i+1} - r_i)}{r_{i+1}} \right| \\
& < \sum_{i=1}^{m-1} \left| \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - \frac{d(r_{i+1} - r_i)}{r_{i+1}} \right| \\
& \quad + \frac{\lambda d}{a} \sum_{i=1}^{m-1} [P_e(r_{i+1})(r_{i+1} - r_i) - P_e(r_i)(r_{i+1} - r_i)]
\end{aligned}$$

where the last inequality follows from the facts that $0 \leq P_e(r_i) \leq P_e(r_{i+1}) \leq 1$ and $r_{i+1} > \frac{a}{\lambda}$. Making use of Lemma 3 and Lemma 4, we have

$$\begin{aligned}
& \left| \sum_{i=1}^{m-1} P_e(r_i) \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - d \sum_{i=1}^{m-1} \frac{P_e(r_{i+1})(r_{i+1} - r_i)}{r_{i+1}} \right| \\
& \leq \frac{d(d-1)}{2} \sum_{i=1}^{m-1} \left(\frac{r_{i+1} - r_i}{r_{i+1}} \right)^2 \\
& \quad + \frac{\lambda d}{a} \sum_{i=1}^{m-1} P_e(r_{i+1})(r_{i+1} - r_i) - \frac{\lambda d}{a} \sum_{i=1}^{m-1} P_e(r_i)(r_{i+1} - r_i) \\
& \leq \frac{d(d-1)\lambda(\lambda-1)\varpi}{2a} + \frac{\lambda d}{a} \sum_{i=1}^{m-1} P_e(r_{i+1})(r_{i+1} - r_i) - \frac{\lambda d}{a} \sum_{i=1}^{m-1} P_e(r_i)(r_{i+1} - r_i).
\end{aligned}$$

□

Lemma 7. For a set of grid points $\mathcal{G} = \{r_\ell \mid 1 \leq \ell \leq m\}$ with $\frac{a}{\lambda} = r_1 < r_2 < \dots < r_m = a$, define function $\aleph(\cdot)$ such that

$$\aleph(\mathcal{G}) = \sum_{\ell=1}^{m-1} P_e(r_\ell) \left[1 - \left(\frac{r_\ell}{r_{\ell+1}} \right)^d \right].$$

Then for any two sets of grid points \mathcal{G}_1 and \mathcal{G}_2 such that $\mathcal{G}_1 \subset \mathcal{G}_2$,

$$\aleph(\mathcal{G}_1) \leq \aleph(\mathcal{G}_2).$$

Proof. Consider two sequences of grid points $\mathcal{G}_1 = \{r_\ell \mid 1 \leq \ell \leq m\}$ and $\mathcal{G}_2 = \{\hat{r}_\ell \mid 1 \leq \ell \leq m+1\}$ such that

$$\frac{a}{\lambda} = r_1 < r_2 < \dots < r_m = a, \quad \frac{a}{\lambda} = \hat{r}_1 < \hat{r}_2 < \dots < \hat{r}_{m+1} = a,$$

and that \mathcal{G}_2 is obtained from \mathcal{G}_1 by adding a grid point \hat{r}_{i+1} to interval (r_i, r_{i+1}) where $1 \leq i \leq m-1$, i.e., $\hat{r}_j = r_j$, $j = 1, \dots, i$ and $\hat{r}_{j+1} =$

r_j , $j = i + 1, \dots, m$. By the definition of function $\aleph(\cdot)$, we have

$$\begin{aligned} & \aleph(\mathcal{G}_2) - \aleph(\mathcal{G}_1) \\ = & \sum_{\tau=1}^m P_e(\widehat{r}_\tau) \left[1 - \left(\frac{\widehat{r}_\tau}{\widehat{r}_{\tau+1}} \right)^d \right] - \sum_{\tau=1}^{m-1} P_e(r_\tau) \left[1 - \left(\frac{r_\tau}{r_{\tau+1}} \right)^d \right] \\ = & P_e(r_i) \left[1 - \left(\frac{r_i}{\widehat{r}_{i+1}} \right)^d \right] + P_e(\widehat{r}_{i+1}) \left[1 - \left(\frac{\widehat{r}_{i+1}}{r_{i+1}} \right)^d \right] - P_e(r_i) \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right]. \end{aligned}$$

By virtue of the fact that $P_e(\widehat{r}_{i+1}) \geq P_e(r_i)$, we have

$$\begin{aligned} & \aleph(\mathcal{G}_2) - \aleph(\mathcal{G}_1) \\ \geq & P_e(r_i) \left[1 - \left(\frac{r_i}{\widehat{r}_{i+1}} \right)^d \right] + P_e(r_i) \left[1 - \left(\frac{\widehat{r}_{i+1}}{r_{i+1}} \right)^d \right] - P_e(r_i) \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] \\ = & P_e(r_i) \left[1 - \left(\frac{r_i}{\widehat{r}_{i+1}} \right)^d - \left(\frac{\widehat{r}_{i+1}}{r_{i+1}} \right)^d + \left(\frac{r_i}{r_{i+1}} \right)^d \right]. \end{aligned}$$

Recall that $r_i < \widehat{r}_{i+1} < r_{i+1}$, we have

$$\begin{aligned} & 1 - \left(\frac{r_i}{\widehat{r}_{i+1}} \right)^d - \left(\frac{\widehat{r}_{i+1}}{r_{i+1}} \right)^d + \left(\frac{r_i}{r_{i+1}} \right)^d \\ = & \frac{\widehat{r}_{i+1}^d - r_i^d}{\widehat{r}_{i+1}^d} - \frac{\widehat{r}_{i+1}^d - r_i^d}{r_{i+1}^d} \\ = & \frac{(\widehat{r}_{i+1}^d - r_i^d)(r_{i+1}^d - \widehat{r}_{i+1}^d)}{\widehat{r}_{i+1}^d r_{i+1}^d} \\ > & 0. \end{aligned}$$

It follows that $\aleph(\mathcal{G}_2) - \aleph(\mathcal{G}_1) \geq 0$. □

We are now in the position to prove statement (II) of the theorem. For any set of grid points, we can reduce the maximal gap between grid points by adding grid points. Every new grid point is placed at the middle of one of the previous intervals which possess the *largest* width in order to ensure that, as more grid points added, the maximal gap of grid points tends to zero. In this process, we create a series of nested sets of grid

points \mathcal{G}_k , $k = 1, 2, \dots, \infty$ such that $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{G}_3 \subset \dots$. Note that

$$\begin{aligned}
 & \left| \sum_{i=1}^{m-1} P_e(r_i) \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - d \int_{\frac{a}{\lambda}}^a \frac{P_e(x)}{x} dx \right| \\
 & \leq \left| \sum_{i=1}^{m-1} P_e(r_i) \left[1 - \left(\frac{r_i}{r_{i+1}} \right)^d \right] - d \sum_{i=1}^{m-1} \frac{P_e(r_{i+1})(r_{i+1} - r_i)}{r_{i+1}} \right| \\
 & \quad + \left| d \sum_{i=1}^{m-1} \frac{P_e(r_{i+1})(r_{i+1} - r_i)}{r_{i+1}} - d \int_{\frac{a}{\lambda}}^a \frac{P_e(x)}{x} dx \right| \\
 (E.2) \quad & \leq \frac{d(d-1)\lambda(\lambda-1)\varpi}{2a} + \frac{\lambda d}{a} \sum_{i=1}^{m-1} P_e(r_{i+1})(r_{i+1} - r_i) \\
 & \quad - \frac{\lambda d}{a} \sum_{i=1}^{m-1} P_e(r_i)(r_{i+1} - r_i) \\
 & \quad + d \left| \sum_{i=1}^{m-1} \frac{P_e(r_{i+1})(r_{i+1} - r_i)}{r_{i+1}} - \int_{\frac{a}{\lambda}}^a \frac{P_e(x)}{x} dx \right|
 \end{aligned}$$

where inequality (E.2) follows from Lemma 6. By Lemma 2, $\mathbb{P}(\cdot)$ is a continuous function with respect to r . Consequently, $\frac{P_e(x)}{x}$ is Riemann integrable over interval $[\frac{a}{\lambda}, a]$ and

$$\lim_{\varpi \rightarrow 0} \sum_{i=1}^{m-1} \frac{P_e(r_{i+1})(r_{i+1} - r_i)}{r_{i+1}} = \int_{\frac{a}{\lambda}}^a \frac{P_e(x)}{x} dx.$$

Moreover, since $P_e(x)$ is Riemann integrable, we have

$$\lim_{\varpi \rightarrow 0} \sum_{i=1}^{m-1} P_e(r_{i+1})(r_{i+1} - r_i) = \lim_{\varpi \rightarrow 0} \sum_{i=1}^{m-1} P_e(r_i)(r_{i+1} - r_i) = \int_{\frac{a}{\lambda}}^a P_e(x) dx.$$

Hence, the right hand side of inequality (E.2) can be made arbitrarily small by successively cutting the gap between grid points in half with new grid points. This proves that

$$\lim_{k \rightarrow \infty} \aleph(\mathcal{G}_k) = d \int_{\frac{a}{\lambda}}^a \frac{P_e(x)}{x} dx.$$

On the other hand, by Lemma 7, we have $\aleph(\mathcal{G}_1) \leq \aleph(\mathcal{G}_2) \leq \aleph(\mathcal{G}_3) \leq \dots$. Combining the convergency and the monotone property of sequence $\{\aleph(\mathcal{G}_k)\}_{k=1}^{\infty}$, we can conclude that $\aleph(\mathcal{G}) \leq d \int_{\frac{a}{\lambda}}^a \frac{P_e(x)}{x} dx$ for any set of grid points \mathcal{G} . By Lemma 5, the expected number of rows of the matrix of violations V^n is no greater than

$$\begin{aligned}
 1 + NP_e(a) + 2N\aleph(\mathcal{G}) & \leq 1 + NP_e(a) + 2Nd \int_{\frac{a}{\lambda}}^a \frac{P_e(x)}{x} dx \\
 & = 1 + NP_e(a) + 2Nd \int_{\frac{a}{h}}^a \frac{P_e(x)}{x} dx
 \end{aligned}$$

for any \mathcal{G} . Such bound applies to any V^j because the number of rows of V^j is non-decreasing with respect to j . Finally, the inequality of (6.4) can be proved by making use of the observation that $P_e(x) \leq P_e(a)$, $\forall x \in [\frac{a}{h}, a]$.

APPENDIX F. PROOF OF THEOREM 6

We need the following lemma, which has recently been obtained in [7].

Lemma 8. *Let X_i , $i = 1, \dots, N$ be i.i.d. Bernoulli random variables such that $\Pr\{X_i = 1\} = 1 - \Pr\{X_i = 0\} = \mathbb{P}_X > 0$. Let $K = \frac{\sum_{i=1}^N X_i}{N}$. Then*

$$\Pr\{\mathcal{L}(K) < \mathbb{P}_X < \mathcal{U}(K)\} > 1 - \delta.$$

Applying Lemma 8, we have $\Pr\{\mathcal{L}(K_{i+1}) < \mathbb{P}(r_{i+1}) < \mathcal{U}(K_{i+1})\} > 1 - \frac{\delta}{2}$ and $\Pr\{\mathcal{L}(K_i) < \mathbb{P}(r_i) < \mathcal{U}(K_i)\} > 1 - \frac{\delta}{2}$. Hence by the Bonferroni's inequality,

$$\Pr\{\mathcal{L}(K_{i+1}) < \mathbb{P}(r_{i+1}) < \mathcal{U}(K_{i+1}), \mathcal{L}(K_i) < \mathbb{P}(r_i) < \mathcal{U}(K_i)\} > 1 - \delta.$$

By the definitions of $\mathbb{P}^*(r)$, $\overline{\mathbb{P}}(r)$ and $\underline{\mathbb{P}}(r)$, we have that event $\{\mathcal{L}(K_{i+1}) < \mathbb{P}(r_{i+1}) < \mathcal{U}(K_{i+1}), \mathcal{L}(K_i) < \mathbb{P}(r_i) < \mathcal{U}(K_i)\}$ implies event $\{\underline{\mathbb{P}}(r) + \varsigma < \mathbb{P}^*(r) < \overline{\mathbb{P}}(r) - \varsigma, \forall r \in [r_i, r_{i+1}]\}$. Hence, $\Pr\{\underline{\mathbb{P}}(r) + \varsigma < \mathbb{P}^*(r) < \overline{\mathbb{P}}(r) - \varsigma, \forall r \in [r_i, r_{i+1}]\} > 1 - \delta$. By Theorem 2 and the gridding scheme, $\Pr\{|\mathbb{P}^*(r) - \mathbb{P}(r)| < \varsigma, \forall r \in [r_i, r_{i+1}]\} = 1$. Applying Bonferroni's inequality, we have

$$(F.1) \quad \Pr\{\underline{\mathbb{P}}(r) + \varsigma < \mathbb{P}^*(r) < \overline{\mathbb{P}}(r) - \varsigma, |\mathbb{P}^*(r) - \mathbb{P}(r)| < \varsigma, \forall r \in [r_i, r_{i+1}]\} > 1 - \delta.$$

Finally, the theorem is proved by observing that the left hand side of inequality (F.1) is no greater than $\Pr\{\underline{\mathbb{P}}(r) < \mathbb{P}(r) < \overline{\mathbb{P}}(r), \forall r \in [r_i, r_{i+1}]\}$.

REFERENCES

- [1] E. W. BAI, R. TEMPO, AND M. FU, "Worst-case properties of the uniform distribution and randomized algorithms for robustness analysis," *Mathematics of Control, Signals and Systems*, vol. 11, pp.183-196, 1998.
- [2] B. R. BARMISH, C. M. LAGOA, AND R. TEMPO, "Radially truncated uniform distributions for probabilistic robustness of control systems," *Proc. of American Control Conference*, pp. 853-857, Albuquerque, New Mexico, June 1997.
- [3] B. R. BARMISH AND C. M. LAGOA, "The uniform distribution: a rigorous justification for its use in robustness analysis," *Mathematics of Control, Signals and Systems*, vol. 10, pp. 203-222, 1997.
- [4] B. R. BARMISH AND P. S. SHCHERBAKOV, "On avoiding vertexization of robustness problems: The approximate feasibility concept," *IEEE Transactions on Automatic Control*, vol. 42, pp. 819-824, 2002.
- [5] X. CHEN, J. ARAVENA, AND K. ZHOU, "Risk analysis in robust control — making the case for probabilistic robust control," *Proc. of American Control Conference*, pp. 1533-1538, Portland, Oregon, June 2005.
- [6] X. CHEN, K. ZHOU, AND J. ARAVENA, "Fast construction of robustness degradation function," *SIAM Journal on Control and Optimization*, vol. 42, pp. 1960-1971, 2004.
- [7] X. CHEN, K. ZHOU AND J. ARAVENA, "Fast universal algorithms for robustness analysis," *Proceedings IEEE Conference on Decision and Control*, pp. 1926-1931, Maui, December 2003.
- [8] H. CHERNOFF, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Annals of Mathematical Statistics*, vol. 23, pp. 493-507, 1952.
- [9] C. J. CLOPPER AND E. S. PEARSON, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, vol. 26, pp. 404-413, 1934.
- [10] W. FELLER, *An Introduction to Probability Theory and Its Applications*, Wiley, 1968.
- [11] P. F. HOKAYEM AND C. T. ABDALLAH, "Quasi-Monte Carlo methods in robust control design," *Proceedings IEEE Conference on Decision and Control*, pp. 2435-2440, Maui, December 2003.

- [12] S. KANEV, B. De SCHUTTER, AND M. VERHAEGEN, "An ellipsoid algorithm for probabilistic robust controller design," *Systems and Control Letters*, vol. 49, pp. 365-375, 2003.
- [13] V. KOLTCHINSKII, C.T. ABDALLAH, M. ARIOLA, P. DORATO, AND D. PANCHENKO, "Improved sample complexity estimates for statistical learning control of uncertain systems," *IEEE Transactions on Automatic Control*, vol. 46, pp. 2383-2388, 2000.
- [14] C. M. LAGO, "Probabilistic enhancement of classic robustness margins: a class of none symmetric distributions," *Proc. of American Control Conference*, pp. 3802-3806, Chicago, Illinois, June 2000.
- [15] C. M. LAGO, X. LI, M. C. MAZZARO, AND M. SZNAIER, "Sampling random transfer functions," *Proceedings IEEE Conference on Decision and Control*, pp. 2429-2434, Maui, December 2003.
- [16] C. MARRISON AND R. F. STENGEL, "Robust control system design using random search and genetic algorithms," *IEEE Transaction on Automatic Control*, vol. 42, pp. 835-839, 1997.
- [17] L. R. RAY AND R. F. STENGEL, "A monte carlo approach to the analysis of control systems robustness," *Automatica*, vol. 3, pp. 229-236, 1993.
- [18] S. R. ROSS AND B. R. BARMISH, "Distributionally robust gain analysis for systems containing complexity," *Proceedings of Conference on Decision and Control*, pp. 5020-5025, Orlando, Florida, December 2001.
- [19] R. F. STENGEL AND L. R. RAY, "Stochastic robustness of linear time-invariant systems," *IEEE Transaction on Automatic Control*, vol. 36, pp. 82-87, 1991.
- [20] Q. WANG AND R. F. STENGEL, "Robust control of nonlinear systems with parametric uncertainty," *Automatica*, vol. 38, pp. 1591-1599, 2002.

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING, LOUISIANA STATE UNIVERSITY,
BATON ROUGE, LA 70803

E-mail address: `chan@ece.lsu.edu`, `kemin@ece.lsu.edu`, `aravena@ece.lsu.edu`

